DOCUMENT RESUME

ED 099 426 95 TM 004 304

AUTHOR	Severy, Lawrence J.
TITLE	Procedures and Issues in the Measurement of Attitudes. TM Report No. 30.
INSTITUTION	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ETS-TM-30
PUB DATE	Dec 74
CONTRACT	OEC-0-70-3797-519
NOTE	13p.
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Attitudes; *Attitude Tests; Scoring; *Test Construction; *Testing; Test Reliability; Test Validity

ABSTRACT

Issues relevant to the nature of attitudes are discussed. The reader is referred to works indexing a variety of existent attitude scales. The way in which one constructs, administers, scores, interprets, and presents findings of an original attitude measuring device is discussed comprehensively, and yet in a nontechnical fashion for administrators, educators, graduate students, novice researchers, and program and project directors. (Author/RC)

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

TM REPORT 30

DECEMBER 1974

PROCEDURES AND ISSUES IN THE MEASUREMENT OF ATTITUDES

Lawrence J. Severy

PREFACE

Many of us assume that the way people feel about something affects their behavior and consequently, the effectiveness of their teaching, the quality of their performance, or the efficiency of a program. "Feelings toward" an object, person, or concept have been labeled "attitudes," and have traditionally been the domain of social psychologists. If there are such things as attitudes, we should be able to measure them. It shall be the purpose of this paper to: discuss issues relevant to the nature of attitudes; refer the reader to works indexing a variety of existent attitude scales; and discuss comprehensively, and yet in a nontechnical fashion for administrators, educators, graduate students, novice researchers, program and project directors, the way in which one constructs, administers, scores, interprets, and presents the findings of an original attitude-measuring device.

Nature of Attitudes

There are two different schools of thought regarding the structural nature of attitudes. The first holds that an attitude is simply the tendency to evaluate an object or construct in positive or negative terms. A definition representative of this position is by Thurstone (1946, p. 39), who suggests that an attitude is "the intensity of positive or negative affect for or against a psychological object. A psychological object is any symbol, person, phrase, slogan, or idea toward which people can differ as regards positive or negative affect." Similarly, Bem (1970, p. 14) suggests that "attitudes are likes and dislikes. They are our affinities for and our aversions to. . . ." It should be clear that this school of thought holds that attitudes are *evaluative*. Secondly, attitudes have an *object*—they refer to something. Writers who conceive of attitudes in this fashion are known as unidimensionalists by virtue of their concentration on one particular dimension, namely evaluativeness. These formulations come closest to the introductory suggestion that attitudes are "feelings toward" something.

A second orientation to the nature of attitudes is provided by a group known as *component* theorists. According to their formulations, attitudes are more than evaluativeness. For example, Wagner (1969, p. 3) suggests that "an attitude is composed of affective, cognitive, and behavioral components that correspond, respectively, to one's evaluations of, knowledge of, and predisposition to act toward the object of the attitude." More comprehensively, Zimbardo and Ebbesen (1970, p. 7) identify the components as follows: "The affective component consists of a person's evaluation of, liking of, or emotional response to some object or person. The cognitive component has been conceptualized as a person's beliefs about, or factual knowledge of, the object or person. The behavioral component involves the person's overt behavior directed toward the object or person." Although component theorization is possibly more comprehensive, it is also more cumbersome. Further, since facts and knowledge may either be more or less enduring than attitudes and behavioral tendencies (as well as overt behavior) may or may not reflect attitudes, the unidimensional approach shall be utilized for the remainder of this paper.

Relationship to Similar Concepts

It is appropriate to distinguish the concept of attitude from others which appear to be similar or related. The list of such terms is probably endless; consequently, only a limited number will be briefly discussed. They are as follows: fact, belief, opinion, motive, mood, personality trait, and temperament. As was mentioned above, there is a difference between fact and attitude. Attitudes tend to be of different duration than facts. The most general argumentation takes the following form: Although we can change the facts regarding certain situations, people tend to evaluate these situations in the same way regardless of the change of facts. For example, one can increase the amount of knowledge a prejudiced person has about blacks without

ever changing his attitudes towards blacks. However, an argument can also be made that attitudes change more quickly than facts. Someone may come to evaluate something more positively even though nothing has changed regarding the factual content of the situation. For example, it is all right to just change your mind.

Traditionally, beliefs have been regarded as one's evaluation of the truth or falsehood of something. Believing that something is true is not the same as thinking positively about it. The term "opinion" is of a similar nature. For example, Aronson (1972, p. 86) claims that "an opinion is what a person believes to be factually true. . . . Compared to opinions, attitudes are extremely difficult to change." In addition to the true-false nature of opinions, Aronson claims that attitudes are more enduring than most other concepts.

There is another set of terms related to attitude as a result of the position that attitudes define a certain readiness to respond in a certain fashion. Although it may make some sense to consider attitudes as motivating some of our behavior, we generally think of attitudes as more enduring than motives. One can speak of a specific intent to achieve a particular goal, and that might be quite different from the person's evaluation or attitude towards that goal. Similarly, the concept of mood is thought to be more momentary. Although we may fluctuate between good and bad moods, attitudes tend to remain a bit more stable.

If one continues with the idea that attitudes are of an enduring nature, one shortly realizes the possibility that attitude is similar to personality traits and/or temperament. Most writers suggest that attitudes are less enduring than these concepts and at the same time, slightly different. For example, we generally think of personality traits as reflecting some typical or characteristic form of behavior of an individual. We call someone "aggressive" who continually displays such behavior in his intentional actions; the attribution of this personality trait depends on the observance of such behavior. Attitudes may not have behavioral referents. Attitudes connotate evaluation or how someone "feels toward" something. The distinction is a fine one as it would be unlikely for someone to be very aggressive if they did not have a positive attitude towards this kind of behavior.

Attitudes and Their Relationship to Behavior

It should be obvious that this formulation does consider attitudes and behavior to be the same thing. According to the position, how someone "feels about" something may or may not be reflected in behavior. Secondly, just because someone behaves toward something in a particular fashion does not really mean that we have measured that person's attitudes. Attitudes are simply one aspect of the behavioral situation. In a vacuum , one's attitudes would lead directly to behavioral characteristics of these attitudes. However, we all recognize that the social-psychological situation (environment) often impinges upon certain behavioral constraints, limitations, and/or prescriptions, which do not always allow for behavior that would be perfectly reflective of a person's attitudes. Consequently, behavior can be thought of as some function of a person's attitudes and other aspects of the particular context of the psychological situation.

Readers interested in empirical analysis of the relationship between attitudes and behavior can be referred to Wicker (1969). Wicker reviews 30 different examples of empirical research analyzing the delineation of the relationship between attitudes and behavior. He draws the inescapable conclusion that feelings are not directly translated into action. Shaw (1973), in a reanalysis of the same data, suggests that only 7 of the 30 studies meet minimal criteria for appropriate measures of both attitudes and behavior. In these seven studies, the demonstrated relationship between attitudes and behavior is substantially higher. Consequently, Shaw's position is similar to that already stated in the preceding paragraph. Attitudes can be expected to lead to a particular kind of behavior given that the situation and other constraints make the behavior appropriate.

General Characteristics of Measurement

As just noted, 23 of the 30 studies reviewed by Wicker did not meet the minimal standards for appropriate attitude assessment. It shall be the purpose of the remainder of this paper to make sure that the reader does not fall into this particular group with his own attempt. The suggested guidelines, if followed by the aforementioned group, might have substantially altered their findings. Before turning to the actual techniques of attitude scale construction, an introduction to general concerns for test adequacy and a discussion of the variety of attitude characteristics that can be measured are in order.

According to Scott (1968), an adequate measure of an attitude would have the following characteristics:

1. It would reflect the intended property veridically.
2. It would be unaffected by irrelevant characteristics, either within the subject or within the testing situation.
3. It would not modify the property in the course of measuring it.
4. It would make sufficiently fine distinctions among persons to represent gradations along the dimension as conceived.
5. It would yield results substantially equivalent to those produced by another adequate instrument measuring the same property.
6. It would yield equivalent scores on a retest administered within a time period in which the property can be assumed to remain constant.
7. It would be relatively easy to construct, administer, score, and interpret.

Scott is, therefore, suggesting that a measure truthfully reflect the attitude and not be affected by extraneous characteristics of the situation. Further, it should not alter the attitudinal characteristic in the process of measuring it, should be as good as any other test measuring the same property, should make fine distinctions, and should, in effect, have utility.

As suggested above, given any particular definition of an attitude, there are different characteristics of attitudes that can be investigated. For example, once one understands what a person is, it is possible to measure and describe different characteristics of that person (e.g., eye color, hair color, height, weight, and so on). What are the characteristics that can be addressed when considering attitudes? Again following Scott (1968), there appear to be approximately eight different characteristics and they are as follows:

Direction. Following the concept of attitudes as "feelings toward," it is clear that a person can have a favorable feeling towards an object or, on the other hand, a negative feeling. Direction merely indicates an individual's tendency to approach, support, and feel positive about an attribute or that subject's tendency to avoid or feel negative about an attribute.

Magnitude. Although direction indicates positive or negative feelings regarding an object, one might still wonder how favorable or unfavorable that attitude is. Direction does not describe the *degree* of favorableness or unfavorableness. Does the subject feel slightly negative or very negative towards a particular attribute?

Intensity. Intensity refers to the "strength of feeling" associated with the attitude. One may feel slightly negative towards a particular attribute, but the evaluation is immaterial if it is not an important issue for him. On the other hand, one might feel slightly negative about an attribute that is very, very important to that person.

Ambivalence. If one thinks of an attitude in bipolar terms, with the direction being either favorable or unfavorable, one can imagine a situation in which subjects have both favorable and unfavorable responses to different aspects of that attitude. The greater the number of these "opposite tendencies," the higher the amount of ambivalence.

Salience/Centrality. The centrality of an attitude refers to the prominence of an attitude. In other words, is this particular attitude an important, focal one by which the individual guides a major proportion of his behavior?

Affective Salience. This refers to how emotional an individual becomes about a particular attitude. We get very emotional about some issues and not about others.

Flexibility. Flexibility connotes the ease with which an attitude can be varied or modified due to persuasive pressure.

Imbeddedness. Whereas some of our attitudes exist in isolation and seem unrelated to other attitudes, other attitudes appear to exist in a network of associations due to the attitude referent's association with a series of other concepts. The degree of isolation versus connectedness can be viewed as the amount of imbeddedness of an attitude characteristic.

According to Scott (1968, p. 208) if one is to "measure attitudes as they are conceptualized in the literature, one needs to find ways of operationalizing, and converting to numbers, such diverse and vague properties as these. In actual practice, most of them have not been operationalized satisfactorily, let alone, scaled. By far, the greatest attention has been devoted to the measurement of magnitude (or intensity). . . ." Although one most often measures the direction and magnitude (and possibly the intensity of an attitude), there are other characteristics that are just as viable for empirical work.

ATTITUDE SCALE RESOURCES

Before constructing an original attitude-measuring device, it is appropriate to seek out information regarding the possibility that there already exist scales of precisely the attribute you wish to measure. There are two reasons that make such an effort worthwhile. The first is obvious—it may save you work. The process of constructing a new attitude scale is an involved one. Second, the existence of such a scale would allow one to compare present findings with previous findings. Although one may develop an ideal scale, it exists in isolation and leaves unanswered questions regarding what attitudes on such a device would be in different settings, in different institutions, under different conditions, or in a previous time in our history. By using a previously constructed scale or a modified version of it, one can refer to earlier work and earlier findings to help place and interpret present findings.

Although one may wish to construct a new scale, the identification of existent scales allows for the incorporation of this older scale with the new effort in its entirety or in partial form. The researcher should not view the situation

as all or none. Do not simply think of using the already existing scale or a new one you would develop. Many situations lend themselves to the utilization of both.

At the time of this writing, there are at least two valuable references containing codified descriptions, interpretations, and evaluations of existing attitude scales: *Scales for the Measurement of Attitudes* by Shaw and Wright (1967) and *Measures of Social-Psychological Attitudes* by Robinson and Shaver (1973). By referring to these works, one is exposed to hundreds of developed scales. The Shaw and Wright work categorizes attitude scales according to the following scheme: social practices, social issues and problems, international issues, abstract concepts, political and religious attitudes, ethnic and national groups, significant others, and social institutions. The Robinson and Shaver volume categorizes: attitudes toward life satisfaction and happiness, self-esteem, internal-external locus of control, alienation and anomie, authoritarianism and dogmatism, socio-political attitudes, values, general attitudes toward people, religious attitudes, and methodological scales.

TRADITIONAL APPROACHES TO SCALE CONSTRUCTION

When a person decides to construct a new scale, decisions need to be made. When someone decides to buy a new car, one must decide what make of car (a Ford, a Cadillac, a Volvo, whatever). There are also different kinds of cars (sedans, station wagons, and so on). Based on what the individual feels familiar with, safe in, or able to afford, a choice is made. Similar concerns face the developer of a new attitudinal scale.

Since there are different types of scales (and, by employing one type, an individual is obviously not using the others), it is perhaps advisable to know the differences between the various approaches so that one can appropriately choose and utilize the type that will be most beneficial to his purposes. Also, by knowing the other types, one can place into context his own efforts. For these reasons, the Thurstone, Guttman, and Likert scales and an Osgood semantic differential approach will be briefly described. The point is that you cannot employ all four methods in any particular scale, and it is important to understand that inherently any scale is qualitatively different from other types of scales. An overview of the major techniques is provided by Zimbardo and Ebbesen (1970, p. 123).

> Each of the techniques to be discussed makes different assumptions about the nature of the test items that are used, and the kind of information they provide about a person's attitudes. However, there are certain basic assumptions which are common to all of these methods. First of all, it is assumed that subjective attitudes can be measured by a quantitative technique, so that each person's opinion can be represented by some numerical score. Secondly, all of these methods assume that a particular test item has the same meaning for all respondents, and thus a given response will be scored identically for everyone making it. Such assumptions may not always be justified but as yet, no measurement technique has been developed which does include them.

The Thurstone, Likert, and Guttman methods of measuring attitudes require subjects to indicate their agreement or disagreement with a series of statements about the object of an attitude. "Generally, these statements attribute to the object characteristics that are positively or negatively evaluated, and rarely neutral" (Shaw and Wright, 1967, p. 13). Therefore, the type of attitude scale developed by these methods measures the acceptance of evaluative statements about the attitude object. Consequently, "the attitude toward the object is inferred from the statements endorsed by the subject, based upon the consensual evaluation of the nature of the characteristics attributed to the object by the acceptance of the statements. Such scales measure only the positivity-negativity of the affective reaction." (Shaw and Wright, 1967, p. 14). It should be clear, then, that these methods reflect the kinds of "feelings toward" that have been described as attitudes. A closer look at the nature of each of these methods follows.

Thurstone's (1928, 1929, and 1931) method of equal appearing intervals was the first major method of attitude measurement to be developed. The Thurstone and Chave (1929) effort described the construction of a measuring device to tap attitudes towards the church. The attempt introduced metric to a virgin area for research. "Thurstone assumed that one could obtain statements of opinion about a particular issue and could order them according to a dimension of expressed favorableness, unfavorableness towards the issue. Furthermore, the ordering of these statements could be such that there appear to be an equal distance between the adjacent statements on the continuum." (Zimbardo and Ebbesen, 1970, pp. 123-124). A unique characteristic of the Thurstone method is that it assumes each statement to be independent of and uncorrelated with the other statements. That is, the acceptance of any one particular statement does not imply the acceptance of another statement. A Thurstone scale is constructed by: 1) formulating and collecting a large number of item statements concerning the object of an attitude; 2) having knowledgeable judges sort the statements into a discrete number of piles or categories (usually 11 categories numbered 1-11 in terms of favorableness) that appear to be equally spaced in terms of degree to which

agreement with the item reflects the underlying attitude; 3) computing a mean score for each item across the different judges; 4) computing a measure of variability for each item; and 5) choosing a set of items for the final scale which have low variability across the judges. In other words, judges need to agree that a particular item represents a particular attitude, and items are chosen so as to represent an even spread across the favorability continuum. A particular person's attitude on an issue is obtained by asking him to check those statements with which he agrees. His final score is the mean scale value of all of the items that he has checked.

Thus, the Thurstone scale consists of a series of statements that are supposedly unrelated. The statements represent intervals along a continuum; the favorableness or unfavorableness of the original items is determined by the judgment of a group of experts regardless of their own attitudes, and a subject's attitude on a particular issue is obtained by noting his responses to the final set of items.

Guttman (1944, 1947) developed another attitude assessment method. His methodology was based upon the assumption that an attitude can be measured by a series of statements which are ordered along a continuum of "difficulty of acceptance." In other words, some of the items in the set should be easy to accept, and others possibly more difficult to agree with. It is assumed that if a person accepts a certain item, that same person accepts all those of a lesser magnitude. Scale items arranged in this fashion are called *cumulative* (if we know the most difficult item a particular subject will accept, we can also predict his attitudes toward other statements). This approach is popular in other testing areas and has been traditionally used in the area of intelligence testing. Guttman's approach is called a scalogram analysis, the essence of which method is to determine whether or not a series of specific items can be appropriately scaled. In other words, the task for someone developing a Guttman scale is to identify a set of items which actually reflects a unidimensional attribute and a cumulative nature as described above. Items which do not fit into this continuum are discarded.

After obtaining a series of statements ranging from very difficult to accept to not at all difficult to accept, a person's attitude is measured by having him check all the statements on the scale that are acceptable to him. His score is determined by examining the pattern of the items he has agreed with.

The type of scale that will be discussed more comprehensively later in this paper is Likert's method of summated ratings. Because the Thurstone scale was somewhat cumbersome and made assumptions regarding the independence of item statements, Likert (1932) developed a technique that could produce an equally reliable attitude scale with relatively less difficulty. The Likert scale is constructed by formulating a series of opinion statements about some issue. Each subject's attitude is measured by asking him to indicate *the extent* of his agreement or disagreement with each statement. Procedurally, this is accomplished by providing each subject with a multipointed scale of response (ranging from strong favorableness to strong unfavorableness). Each person's attitude is then obtained by summing the individual ratings on the different items.

This method assumes that all of the statements reflect the same attitudinal dimension and are therefore related to each other (unlike the Thurstone assumption that items are independent and not related). Furthermore, the Likert approach does not assume equal intervals between the scale values. Consequently, as Zimbardo and Ebbesen (1970, p. 126) point out, "this means that a Likert scale can provide information on the ordering of people's attitudes on the continuum, but is unable to indicate how close or how far apart different attitudes might be." As is true with most other scales, the final scale is composed of those items which best distinguish between subjects with the highest and lowest total scores and which, in turn, distinguish between criterion groups on the attitude.

The Osgood semantic differential (Osgood and Suci, 1955; Osgood *et al.*, 1957) may not really be a method for constructing an attitude scale per se, but rather a way of measuring attitudes. Whereas the Likert, Thurstone, and Guttman scales require subjects to indicate the degree of their agreement with a set of items reflecting an attitude, the semantic differential asks subjects to rate a particular attitude object on a series of bipolar semantic scales. For example, a subject's attitude toward research might be measured by his ratings on a set of bipolar adjectives such as: good-bad, strong-weak, fast-slow, active-passive, each with seven data points between.

Osgood has demonstrated that three general factors of meaning are measured by the semantic differential technique - an evaluative factor, a potency factor, and an activity factor. Since we have been discussing attitudes as being evaluative, the evaluative factor would seemingly measure both the direction and intensity of an individual's attitude toward the object being rated. Further, the evaluative factor seems to be the most important factor or aspect of meaning as measured by the semantic differential. "The bipolar scales having high loadings on this factor [evaluativeness] are good-bad, beautiful-ugly, sweet-sour, clean-dirty, tasty-distasteful, valuable-worthless, kind-cruel, pleasant-unpleasant, bitter-sweet, happy-sad, sacred-profane, nice-awful, fragrant-foul, honest-dishonest, and fair-unfair. In actual practice, the number of bipolar items used varies from all of the fifteen listed above to a few (3 to 5) of the most clearly evaluative pairs. For greater reliability, the attitude score may be computed as the sum or average of the ratings of all scales used" (Shaw and Wright, 1967, p. 30). Consequently, by placing the attitude object at the top of a series of such bipolar adjective scales, one can measure the extent of a person's attitude toward that object.

By way of summation, four different and rather well known methodologies have been employed to tap attitudes. Others do exist, and still others will probably be developed; however, these are the more popular procedures. These techniques involve different assumptions and different demands on time and effort. By employing one methodology, a researcher accepts the particular assumptions of that approach and should be cognizant of the qualitatively different approaches that may be utilized by others attempting to devise scales measuring similar attitudinal objects.

NEW SCALE CONSTRUCTION AND APPLICATION

Scale Construction

As previously identified, Likert's scale appears to be the most popular in present research. Simplistically stated, the goal of this approach is to generate a series of statements or items which reflect the subject's opinion regarding the attitude object in question. A subject is provided a response continuum ranging from favorable to unfavorable, and the researcher simply adds the score on each item to obtain a cumulative total which indicates the subject's attitude toward the object in question. With this overview in mind, it is now appropriate to turn to a more detailed description of the procedure.

Item Selection. The essence of the Likert approach is to provide the respondent with a well-thought-out series of statements that will as accurately as possible reflect the attitude in question. There are a number of concerns one should be aware of during this phase of scale construction. To begin with, the items should, as far as possible, reflect the attitude in question rather than be tangentially related. Further, they should be more representative of the attitude object being studied than any other attitude object. Given these two guidelines, there are a couple of ways of generating items. First and foremost, the scale developer should have a well-thought-out conceptualization of the nature of the attribute that he is attempting to measure. This conceptualization can spring from a theoretical foundation, from a practical knowledge of the situation, or from interaction with other experts regarding that attitude. If, for example, one wanted to investigate an attitude towards helping others, one could start with the premise that helping in groups is different from individual helping and that task helping is different from psychological help. Given these two dichotomies, four different situations exist. For a scale to comprehensively assess helping attitudes, one would want to generate items reflective of each of the four situations. This type of approach has been called the logical, rational, or conceptual approach to item generation.

The empirical approach, on the other hand, suggests that it does not matter where an item comes from, what it sounds like, or if it is theoretically related to the attitude. Scientifically, it is a fine item if it can successfully discriminate between the groups one wants to discriminate between. For instance, the item "I like cold weather" may not theoretically be related to attitudes toward population control. However, if it is known empirically that persons in favor of population control always respond in a fashion different from those opposed to it, then the item is a good one for an attitude scale regarding attitudes toward population control.

Another possibility exists for generating items. As suggested earlier in this paper, it is appropriate to search out previous work in the area before attempting the construction of a new attitude scale. Often you cannot find exactly the measure you are interested in having but, rather, something closely related or something partially correct. In such cases, it is beneficial to use the items from published work that are reflective of the attitude you are attempting to measure. In other words, items from already existing scales can be combined with your own original items to help generate the total set you will begin working with. Given that you will be attempting to measure a single attitude, it is appropriate to start with a total set or item pool consisting of between 30 and 50 items. As suggested, these items can be generated in a variety of ways, but the worth of the entire scale depends most crucially on the appropriate and well-founded choice of the items.

Some guidelines for the wording of items is necessary. One should always be cognizant of the sophistication or literacy level of the population one wishes to work with. If the scale is to be utilized for one specific group, it can be aimed for that group. If it is intended for wide use, then one need be very concerned about keeping the items in a mode that can be easily read and easily understood. A related point is the complexity of any particular item. Remember that a Likert scale attempts to provide information regarding a unidimensional concept. Consequently, each item should be unidimensional. For example, the item "I don't like to feel crowded because I am a very nervous person" would be a bad item for an attitude scale: A subject could agree or disagree with either phrase in the item. Such double-barreled items create a situation wherein the researcher does not know which aspect of the item is being responded to. In this case, it would be much better to create two separate items; one dealing with the response to crowdedness, the other dealing with how nervous the individual may be. The important point here is that many item writers attempt items which are too complex. They attempt to "explain" the behavior in the item. The point is that in

an attitude scale, you are *not* interested in the "whys," you are interested in the attitude itself, favorable or unfavorable.

Items cannot be too simplistic. Items such as "I like to be around pleasant people" may not be very good for a variety of reasons. First, everybody would probably agree with the statement. Since the thrust of attitude measurement is to differentiate among people regarding an attitude, you do not want an item with no variability. More technically, each item must have a fair degree of variance. Items such as this one could be altered slightly to create a bit more variability by adding qualifiers. Another possibility is to word the item in a negative direction, such as "Sometimes I don't like to be around pleasant people." An item should create variability in response, and the easiest way of insuring that is to try to construct items that you would guess half the people in your sample would agree with and half would disagree with. This would generate a mean score for an item at the midpoint of the response continuum. (This is generally known as creating equal probabilities of passing or failing any particular item, which is desired.)

As previously mentioned, items can be worded in both positive and negative directions. Further, there is good reason to attempt a counterbalance–half of the items written in a favorable direction and half written in an unfavorable direction. Often respondents to attitude scales develop what is known as a response or acquiescence set. Under such conditions, the subject does not really attend as precisely as possible to each of the items he is reading. It is easy to slip into this response mode if the subject is allowed to simply agree with all of the items. By reverse wording some of the items, the subject must disagree to remain consistent in his responses. For example, a positive item such as "I feel I am the master of my own fate" could be combined with an item such as "I don't feel like I am responsible for what happens to me." This forces a consistent respondent to use both ends of the response continuum. Briefly then, it is wise to attempt to include an equal number of positively and negatively worded items. A further point concerns the ordering of these items. Since it can be demonstrated that certain items will affect the response to latter items (unless the researcher has the potential for counterbalancing the order of presentation with different subjects), the best approach is that of randomization of the items. Therefore, after generating the original pool of items, a table of random numbers should be used to order them. One last check should be made to insure that not more than four or five positively or negatively worded items occur in sequence. If this is the case, the order of a couple of items should be altered.

It is rather hard to write items without thinking of the response categories or the response continuum that is going to be provided for the subject. Several concerns are relevant here. Remember that you are going to ask subjects to either agree or disagree with the statement you have generated. This does not mean, however, that only two response categories are possible. Researchers generally provide between three and seven response possibilities. Obviously, the more categories you provide, the more sensitive you are asking the subject to be regarding each particular statement. If you have a small number of statements in your scale, a larger number of response categories may be appropriate. If, on the other hand, you have a large number of items, three categories may be sufficient. Consequently, the choice of three, five, seven, or nine response categories depends on the nature and number of the items in your scale. Standard response continuums for a) three, b) five, seven, and d) nine are as follows: a) agree, neither agree nor disagree, disagree; b) strongly agree, agree, neither agree nor disagree, disagree, strongly disagree; c) strongly agree, agree, slightly agree, neither agree nor disagree, slightly disagree, disagree, strongly disagree; and d) very strongly agree, strongly agree, agree, slightly agree, neither agree nor disagree, slightly disagree, disagree, strongly disagree, or very strongly disagree.

It should be pointed out that the above response continuums all include an odd number of items, allowing therefore for a central category of neither agree nor disagree. Many researchers wish to avoid allowing the subject such a response and, alternatively, opt for an even number of categories. This has the effect of forcing the subject towards the positive or negative side. For example, consider the six-item response continuum of strongly agree, agree, slightly agree, slightly disagree, disagree, or strongly disagree. Depending again on the nature of your items, this may be a viable alternative. It does, however, exclude the possibility of a subject indicating a truly ambivalent or midpoint response.

Although the agree versus disagree response continuum is most commonly utilized, other bipolar continuums are just as viable given the nature of the items and thereby the nature of the attitude being assessed. Items can be written so as to reflect many of the characteristics presented earlier. For example, you might ask subjects to respond to items in terms of whether or not the items are: very important, slightly important, neither important nor unimportant, unimportant, or very unimportant. Alternatively, it is possible to ask how *interested* the subject is in the item statements presented. Another popular approach requires respondents to judge the degree of *truthfulness* of the items. In fact, any of the bipolar continuums discussed by Osgood in his work on the semantic differential would be appropriate.

Although we have been primarily concerned with the evaluative dimension in the above response continuums, it is also possible to address the aspects of certainty and salience. For example, for the item "When you think of the future realistically, how certain are you that your economic situation will be better than that of your parents?" a response continuum such as very sure, sure, neither sure nor unsure, unsure, or very unsure would be appropriate. With

regard to salience, the item "How important is a clean environment to you?" lends itself to a response continuum of care very much, care, care a little, or care not at all. Thus, a variety of response continua is possible depending on the nature of the characteristic that is to be measured. The guidelines are those previously mentioned; namely, create a situation in which there is variability on the response continuum and allow for a situation in which half the people respond positively and half the people respond negatively to the item.

Initial Data Collection. After the original pool of items has been written and fitted to an appropriate response continuum, the order randomized, and approximately half the items positively worded and half negatively worded, the scale developer must make arrangements to have a sample of respondents complete the attitude-measuring device on its initial form. Care should be taken to use a group similar in nature to the group for which the final research was designed. (A few of the procedures in upcoming sections on reliability and validity may be attempted at this juncture. Also procedures to be discussed in the section on scale administration should be closely adhered to.) On the basis of this initial administration, which should involve roughly twice as many subjects as there are items in the pool, items can be excluded from the final scale or refined according to the guidelines already discussed in this section—namely, all items not having a substantial amount of variability throughout the response continuum should be discarded, and all items for which the mean or typical response occurs toward the ends of the response continuum should either be discarded or reworded so as to create a more desirable result.

Reliability and Internal Consistency. One of the two major concerns regarding measurement is the reliability of the measuring device. In common language, an attitude scale is judged to be *reliable* when the scale provides the same score for the same subject, given that there is no reason to assume the subject's attitude has changed. For example, we think a ruler is a good measuring device if it provides the same measurement for the same block of wood and changes only when we cut the piece of wood in half. There have been different approaches to the assessment of reliability in attitude scales, and a few of them will be briefly discussed.

One of the most common forms of reliability obtained on scales is known as *test-retest reliability.* The essence of this approach is to ascertain whether or not subjects will score the same at two different points in time. (Again, given there is no reason for you to assume that a change has taken place in people's attitudes.) In keeping with the orientation of this paper as being nontechnical as possible, the test-retest approach has many benefits. If you are completely unskilled at statistics, you simply check to see whether or not the subjects' scores for the first administration are quite similar to the second administration. If the scores seem to be varying widely, your scale is not reliable, and therefore, it should be rejected. If, on the other hand, scores are highly consistent, then you probably do have a reliable scale. For those having statistical sophistication, a correlation coefficient can be run between results obtained with the first administration and the second administration; and for those who are even more skilled, formulas for more exact statistical tests of reliability can be found in Edwards (1970), Cronbach (1960), and Scott (1968). Here again, the higher the correlation, the greater the probability that the scale is reliable in a test-retest sense.

Another form of reliability which can be obtained without a great deal of statistical sophistication is known as *split-half reliability.* In essence, the thrust of the split-half methodology is to randomly assign half of the items of the original scale to a fictitious "form A" and the other half to a fictitious "form B." Subsequently, scores for form A are compared with scores for form B (of course, one must make sure that an equal number of positively and negatively worded items are included in each form of the scale). The higher the degree of comparability between the two forms of the scale, the higher thy potential for reliability. If there is no correspondence between the scores on the two forms, then reliability is probably low. Again, scale developers with more statistical sophistication should consult the previously mentioned references for appropriate formulas.

It is probably appropriate to mention three rather well known statistics discussed under the heading of reliability of scales. Coefficient *alpha* and Kuder-Richardson's (Formula 20) coefficient give statements of reliability. They estimate the degree to which your test will correlate with any equivalent test of the same attribute. Formulas for both of these statistics can be found in Edwards (1970). It should be pointed out, however, that as is true with most forms of reliability, estimates of reliability as demonstrated by these formulas are dependent upon the length of the scale (the number of items you have in your scale) and the homogeneity of the items (the average interitem correlation). Consequently, one way to increase your potential for having a higher *alpha*, or higher reliability of any scale, is to include a larger number of appropriate items.

The average interitem correlation of the items is very important. The extent to which item responses are intercorrelated provides a measure of the *internal consistency* of the items in a scale. Obviously, you would not want all items to intercorrelate perfectly because then you would have nothing more than a large number of measures of exactly the same thing. (That would be redundant, and you would only need one measure.) On the other hand, you do not want your items to be totally uncorrelated as they would probably be measuring different things. Consequently, as is true with most things, you want items which are fairly well related, but not too highly inter-

related. Measures of appropriate internal consistency are available, and Scott's homogeneity ratio (HR) represents the average level of interitem correlation. It is "equal to a weighted average interitem correlation in which the correlation between every pair of items is weighted by the geometric of their variances" (Scott, 1968, p. 254). Generally speaking, if you have made sure that there are equal probabilities of the passing and failing of each of your items, you will have good variances, and if you desire a scale which has the best chance of placing a subject anywhere along the scale continuum, then you desire a homogeneity ratio of .33. If you are statistically unsophisticated, you might simply look at a correlation matrix of your items and determine if the items in your scale tend to be correlated at approximately this level.

One last technique is commonly utilized to insure the internal consistency of the scale Each item in your scale should be correlated to the total score for that scale. Any item which does not correlate positively to the total score in a significant fashion should be excluded from the scale.

If you have appropriately worked through the item-generation phase, have correctly weeded out those items with poor variability and inappropriate means, have obtained sufficient information to convince you that your scale is internally consistent and reliable, then you should proceed to the next stage of scale construction known as validation.

Validity. Whereas reliability is concerned with the repeatability of a particular measurement, *validity* is generally concerned with the truth of a particular measurement. Although I may have a ruler that consistently, reliably tells me a football field is 140 yards long, either the football field is constructed incorrectly, or my ruler is invalid since most football fields are 100 yards long (not including end zones). To repeat, then, the concern for validity means concern for whether or not this attitude scale you have developed actually measures the attitude that you are hoping to tap. It has been traditionally proposed that the validity of a scale is delimited by the reliability of a scale. If reliability is low, then there is little chance that there will be any validity. There are several forms of validity, and we shall briefly describe a few that may be used for those not adept in statistics.

A very popular way of obtaining validity for attitude scales is to work with *criterion groups*. Suppose that you are working with attitudes toward population control. One would expect that members of Zero Population Growth would have different attitudes on your scale than would "Right to Life" organizations. By administering your scale to these two different organizations, you would expect to find very different responses for the two groups. If, in fact, you do get different responses, then an indication of validity is obtained.

There is, however, the possibility that your scale measures something else which also differentiates the two groups (e.g., religiosity). Consequently, there is generally a concern for what is known as *construct validity*. The concept of construct validity emanates from the work of Campbell and Fiske (1959). They suggest that one comprehensively consider the construct or attitude object one is attempting to measure, and attempt to delineate conceptually, theoretically, and so on, the nature of the relationship that construct has with a set of other constructs. After such analysis, attempts to empirically measure the relationships between these constructs should be undertaken with your attitude scale and appropriate taps of the other constructs. For example, continuing on with the concern for population attitude scale, one would probably come to the conclusion that such an attitude would be negatively related to traditional religiosity, positively related to concern for the environment, and possibly not at all related to attitudes towards Volvos. If this thinking is appropriate and correct, then one might attempt to obtain *convergent validity* by demonstrating a positive correlation between the newly developed measure of attitudes toward population control and attitudes towards the environment as measured on a previously existing scale, and *discriminate validity*, by demonstrating a negative correlation between traditional religiosity and your new scale. To complete the picture of construct validity, one might also be able to demonstrate no relationship between attitudes toward Volvos and your new scale.

Another form of validity that is desirable to obtain is that of *predictive validity*. The essence of predictive validity is to group your subjects on the basis of their responses to your scale, and then predict for the different groups differences in relevant behavior. For example, if you have developed a new population-control scale and have obtained responses from 200 individuals, you might take the top 20 and bottom 20 scores on the scale and then predict differences in their contraceptive behavior. If you could demonstrate that such behavior was different for the two groups, then you would have an example of predictive validity.

If you have followed the preceding phases appropriately, and have obtained a set of items which generates a reliable and valid instrument, you now have a scale ready for use.

Administration

There are a number of guidelines regarding the administration of your attitude scale, whether for scale construction or actual use. It has already been mentioned that it is desirable to be quite serious about cohort, or respondent, selection. If you are attempting to develop an attitude scale that is sensitive to the concerns and attitudes of teachers, it is advisable to utilize representative groups of teachers throughout the entire construction process. If you will be working with subjects who are mentally disadvantaged or

are currently institutionalized, then subjects should be chosen from similar circumstances during scale development. Above and beyond these suggestions, which would help insure that the level of your items can be handled by your subjects in the final utilization of the scale, one should also be concerned with the ethics of attitude assessment.

As is true with all research in the social sciences, every effort should be made to insure that the rights of the subject are not violated. This is a truism whether your subjects are teachers, administrators, students, or residents of a total institution. Further, whenever possible, research should be designed so as to make it unnecessary to identify individual subjects and their responses. For example, if you are concerned with the relationship of your measure to a few others, and the nature of those attitudes in juxtaposition to a similar program elsewhere in the country, you do not really need the names of the individuals on your protocols. Simply have subjects place some identifying characteristic on each of the protocols they fill out, and mark the name of the program or institution at the top. In this way, individual responses can be kept anonymous while probably insuring accurate and truthful information.

The concern for obtaining the most accurate information possible from subjects and the avoidance of fabrication should be a prime concern of anyone using attitude-assessment devices. Every effort should be taken throughout the investigative phase to establish a firm rapport with the subjects as well as anyone else involved with your project. For example, if you are working in a total institution, it is best not to create any problems with administration, staff, or the residents themselves if you wish to obtain accurate information from any of the three groups, as they do interact.

During the actual administration of your attitude scales, every effort should be made to limit the length of your battery of attitude scales so as to keep the subject attentive. A general rule of thumb for this is approximately 30 minutes. The size of the group you work with at any one point in time should be determined on the basis of how well you can control the situation. If you need to help a large number of your respondents with the reading of particular items, you will want to keep the group small. If you perceive no such problems but perceive a potential for disruption due to social interaction on the part of group members, it is again best to work with a few subjects at a time. If, on the other hand, you can keep the attention of your respondents on the task and do not feel that subjects will discuss the answers, then you can work with larger groups.

In introducing the scale to your respondents, every effort should be made to appear as straightforward and open as is possible. If it will not contaminate your results, describe exactly what you are about, why you need their help (emphasize the fact that you are asking for their help), and that you would like people to fill out the scales as precisely, as accurately, and as honestly as they possibly can. Emphasize the fact that your work will be confidential and that you will not attempt to identify subjects personally. In other words, make every attempt you can to be open and to model the potential for accurate and open statements of subjects' attitudes.

A final grade for attitude-scale administration has to do with the concern for standardization. It is very important that all subjects be "tested" in similar conditions. You do not want differences in the room, in the lighting, in the mood of the subject or the examiner, in the instructions, or in interpretations of items to account for differences in your derived attitudes. Consequently, if someone raises his hand and asks for an interpretation of a particular item, you should simply restate the item as it is worded on the page. Attempts to use other language or explain the item only serve to give that respondent more information or a different kind of information than was available to all others. With the above in mind, it behooves those working with attitude scales to comprehensively plan for the administration of their scales. A detailed and standardized approach to the situation, the instructional set, and the explanation of any particular item should be worked out well in advance of actual administration. All of these efforts help to create a potential for truthful information.

Scoring

The ease of scoring a Likert scale is one of its most appealing characteristics. It is a cumulative scale; therefore, one simply totals the score of the responses to the items in the scale. Although this sounds easy enough, there is some complexity. Remember that approximately half of your items are positively worded and the other half negatively worded. Secondly, you have chosen one of a variety of different response continua. Yet, the procedure for working through these complexities is not difficult. Start by assigning a scale value to each of your responses. If you are using a five-point continuum of strongly disagree to strongly agree, let strongly disagree stand for 1 point; disagree, 2 points; neither agree nor disagree, 3 points; agree, 4 points; and strongly agree, 5 points. Similarly, 7 points can be assigned to seven-point scales, 9 for 9, 6 for 6, and so on. After assigning scale points to each of the responses, determine which items in the scale are positively worded and which are negatively worded. For the positive items, add up all of the scale values; save this subtotal labeled "A." Treat the negative items in the following fashion. First, count the number of negative items in your scale. Multiply the number of items by the number of categories in your response continuum, plus 1 (e.g., for the five-point continuum, the number would be 6). The product of this multiplication then has subtracted from it the score on each one of the negative items. The resultant subtotal ("B") is then added to the score for the positive

items ("A"), and the total score for the scale is derived.*

Consider the following example. You have constructed an abbreviated four-item scale with two positive items and two negative items. You have provided your subjects with a five-point response continuum, and the subject has strongly agreed with the first item, disagreed with the second item, agreed with the third, and strongly disagreed with the fourth. You know that the first and third items are positively worded. You add the score for the two positive items together (5+4) for a subtotal (A) of 9 points. Since your response continuum is five, you add 1 making 6, and multiply that by 2 (since you have two negative items) for a total of 12. You now subtract from 12 the scores for the second and fourth items (2 and 1), leaving yourself with a subtotal (B) of 9. Adding the two subtotals together, your scale indicates that this particular subject has a total attitude-scale score of 18 points.

Utilization

There are several aspects regarding the utilization of any particular attitude scale. First, there is the interpretation of your data and second, presentation of the information you have obtained. According to the basic assumptions and the procedures involved in the development of a Likert scale, interpretations of responses to such a scale should be confined to the ordering of subjects on the attitude rather than discussions about how large the differences in attitudes may be. More specifically, Likert scales provide ordinal information. This does, however, allow for a considerable amount of utility.

The first consideration involves whether the responses to the new attitude scale should be considered as independent or dependent variables. If you are attempting to demonstrate attitudinal differences between religious groups, between administrators and staff members, between students in one school versus another, then the attitudinal scores are your dependent measures. In such instances, you want to make statements about a different attitude that exists or is created in one situation versus another situation. Ordinal scales do allow you to say that one has a more favorable attitude towards the attitude object versus a less favorable attitude. Recent information of this type has been popular in political opinion polls regarding U.S. citizen's attitudes towards the Nixon resignation. Treating such attitudes as dependent variables, one might, for instance, note that favorableness of resignation was higher among Democrats than Republicans.

Another way of utilizing the scores on the attitude scales is to consider them as the independent variables. In this case, one might simply be interested in how many low scores versus high scores there are. For example, how many people had attitudes that were favorable towards Nixon's resignation versus how many had attitudes unfavorable to his resignation? According to an issue of *Newsweek* (1974), 79 percent of those interviewed were favorable to the resignation versus 21 percent opposed. Another way of utilizing the attitudinal score as an independent measure is to use the attitude measure to separate two different groups for further study. For example, you might wish to look at personality differences between your high and low scorers. Continuing on with our example, you might be interested in describing the personality differences between individuals who had favorable attitudes toward Nixon's resignation versus individuals who had unfavorable attitudes toward his resignation.

Consequently, above and beyond being able to look at scale responses and knowing that those people who score a lot of points are more favorable to the attitude then those who score a few points, you can utilize the information from your scale in a variety of fashions. The most appropriate nonsophisticated statistics to employ would be: correlations (correlations between your scales and other scales, or behavioral measures), and t-tests (t-tests between low scorers and high scorers on your scale, on some other attribute or behavior, or t-tests between two groups on the scores as derived from your scale).

The presentation of the development (and findings) of a particular scale necessarily involves the delineation presented in this paper. Consequently, one should describe where the attitudinal concern comes from, the way items were generated, item refinement, the choice of subjects, reliability work, validity work, and then the actual utilization of the scale in a meaningful arena and the results so obtained. By way of recapitulating the entire process (and in order to serve as a model), this paper concludes with sample presentation sections from previously developed scales.

*This procedure is identical to reverse coding the response continuum on the negative items.

EXAMPLES OF PRESENTATION

In this section of this paper, an abbreviated presentation of an attitude scale is pieced together from three different scales. The first is an attitude scale measuring staff and patient attitudes toward mental health treatment (Swanson and Severy, 1970). The second is a scale measuring attitudes toward population control (McCutcheon, 1974), and the third is a scale tapping individual differences in helping dispositions (Severy, 1975). Clearly, each one of these sections would be more extensive in an actual write-up of the scale development.

Introduction

Previous studies have indicated "that opinions about the adequacy of a mental hospital and staff are useful for the understanding of a patient's response to treatment. Further, the evidence supports the practicality of administering questionnaires to staff members as well as to patients, and the need for a multiple scale instrument to measure various aspects of mental treatment. However, previous scales have examined mental hospitals only as undifferentiated entities or . . . without regard to an evaluative component. Research was designed to develop instruments measuring several aspects of a mental hospital setting" (Swanson and Severy, 1970, p. 80).

Method

Subjects. "Data were collected initially on 286 college students, and 75 non college community members. As a result of these administrations, a second version was constructed which was employed in a study in which 120 college students participated . . . when the last refinement was completed, and the item pool was administered to 135 college students. The four studies represent a total subject pool of 611 subjects" (Severy, 1975).

Item Pool Generation. "A seven choice Likert type attitude scale was developed as a result of an attempt to deal with a wide range of issues related to population control . . . the choices ranged from agree very strongly to disagree very strongly, the preliminary scale had 41 items, about equally divided with respect to the expression of pro and con viewpoints . . . 11 of the items either discriminated poorly or not at all between high and low scores. These items were dropped . . . the final version of the scale included 30 items . . . so that agreement with them is indicative of unfavorable attitudes toward population control" (McCutcheon, 1974, p. 1236).

Results

Scale Reliability. "The reliabilities of the perceived general staff orientation and the adequacy of treatment scales were considered adequate (*alphas* of .81 and .77 respectively). The perceived staff awareness of patients and the perceived staff agreement scales were marginally adequate (*alphas* of .61 and .56 respectively)" (Swanson and Severy, 1970, p. 85). "Odd-even, split half reliability for 131 randomly selected subjects was .89 . . . using the Spearman-Brown formula. Test-retest reliability for 96 subjects was .91" (McCutcheon, 1974, p. 1239).

Validity. In an attempt to get some type of criterion group validity, "ninety-two students from the same community college were first asked to fill out the attitude scale (population opinion). Twelve to fourteen days later they were given a questionnaire and asked to select between two and five topic areas within psychology that they considered to be most important. Included among the list of 16 topics was that of psychological effects of over-population. . . . Twenty-one of these students chose psychological effects of over-population as one of the most important topic areas. Attitude scores of these subjects were compared with those of a randomly selected of 35 subjects chosen from among those who did not select psychological effects of over-population as a topic area. A point bi-serial correlation of .95 ($df=54$, $p. < .001$) was obtained between attitude scores and choice of psychological effects of over-population as an important topic for classroom readings" (McCutcheon, 1974, p. 1239).

"With regard to convergent-divergent validation, recall that it was predicted that the 'need to help people' subscale would correlate higher with our composite scales than would the 'need for people' which would also be positively related. The interrelationship between Harvey's measure of 'anomie' and helping scores would be minimal, and lastly, the measure of 'interpersonal aggression' should be negatively correlated with our helping measures. The total pattern of these intercorrelations is exactly what was desired. Correlation with the helping dispositions total score with Harvey's 'need to help people' was .70, with the 'need for people,' .54, and with 'anomie,' −.08, and with 'interpersonal aggression,' −.17. The pattern of these interrelations provides further evidence to the strength of these scales" (Severy, 1975).

Survey Results (Attitude Scale Results). A fictitious example of one method of presenting attitude scale results can be derived from the *Newsweek*-Gallop poll survey conducted immediately after former President Nixon's resignation. Consider the attitude of favorableness towards resignation. The following table format would be appropriate.

Resignation Favorability

Resignation	*Percentage*
Strongly Favor Resignation	48
Favor Resignation	31
Do Not Favor Resignation	5
Strongly Do Not Favor Resignation	8

Discussion

"The reliability of the new scale appears to have been adequately demonstrated . . . both split-half and test-retest reliability coefficients fell well within the range of acceptability (McCutcheon, 1974, p. 1240). ". . . these measures of construct validity seem to demonstrate the construct validity of the new scale . . . most strong, our relationships

between the scale and the number of children expected, social attitude scale, and birth control scale" (McCutcheon, 1974, p. 1241).

"Our results have shown: a) that such measures can be constructed, b) that the questionnaire approach is feasible for both staff and patients, and c) that these scales may be useful for understanding treatment effectiveness. In contrast to earlier work we provided both staff and patient items . . . and they have demonstrated an ability . . . work is now under way to expand these scales in order to increase reliability and clarity of the measures. These scales should then provide an additional tool for the prediction of treatment effectiveness and the diagnosis of treatment-unit difficulties" (Swanson and Severy, 1970, p. 90).

"It appears that an internally consistent instrument has been developed which is related to real helping behavior . . . the overall consistency of positive findings with regard to the total composite should give credence to the programatic attack and the conceptualization of helping behavior presented above. . ." (Severy, 1975).

REFERENCES

Aronson, E. *The social animal.* San Francisco: W.H. Freeman and Co., 1972.

Bem, D.J. *Beliefs, attitudes, and human affairs.* Belmont, Calif.: Brooks/Cole. 1970.

Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin.* 1959, 56, 81-105.

Cronbach, L.J. *Essentials of psychological testing.* New York: Harper and Row, 1960.

Edwards, A.L. *The measurement of personality traits by scales and inventories.* New York: Holt, Rinehart and Winston, 1970.

Guttman, L. A basis for scaling qualitative data. *American Sociological Review*, 1944, 9, 139-150.

Guttman, L. The Cornell technique for scale and intensity analysis. *Educational and Psychological Measurement*, 1947, 7, 247-280.

Lemon, N. *Attitudes and their measurement.* New York: Wiley, 1974.

Likert, R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932, 140, 1-55.

McCutcheon, L.E. Development and validation of a scale to measure attitude toward population control. *Psychological Reports*, 1974, 34, 235-242.

Newsweek, August 19, 1974, 84(8), 13-20.

Osgood, C.E., & Suci, G.J. Factor analysis of meaning. *Journal of Experimental Psychology*, 1955, 50, 325-338.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. *The measurement of meaning.* Urbana, Ill.: University of Illinois Press, 1957.

Robinson, J.P., & Shaver, P.R. *Measures of social-psychological attitudes.* Ann Arbor, Michigan: Institute for Social Research, the University of Michigan, 1973.

Scott, W.A. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. II. Reading: Addison-Wesley, 1968. Pp. 204-273.

Severy, L.J. Individual differences in helping dispositions. *Journal of Personality Assessment*, 1975, 39, in press.

Shaw M.E. A theory of attitudes. Unpublished manuscript, University of Florida, 1973.

Shaw, M.E., & Wright, J.M. *Scales for the measurement of attitudes.* New York: McGraw-Hill, 1967.

Swanson, R.M., & Severy, L.J. Measuring staff and patient attitudes toward mental health treatment. *Journal of the Fort Logan Mental Center*, 1970, 6, 79-91.

Thurstone, L.L. Attitudes can be measured. *American Journal of Sociology*, 1928, 33, 529-544.

Thurstone, L.L. The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 1931, 26, 249-269.

Thurstone, L.L. Comment. *American Journal of Sociology*, 1946, 52, 39-40.

Thurstone, L.L., & Chave, E.J. *The measurement of attitude.* Chicago: University of Chicago Press, 1929.

Wagner, R.V. The study of attitude change: An introduction. In R.V. Wagner, & J.J. Sherwood (Eds.), *The study of attitude change.* Belmont, Calif.: Brooke/Cole, 1969. Pp. 1-18.

Wicker, W.A. Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social issues*, 1969, 25, 41-78.

Zimbardo, P., & Ebbesen, E.B. *Influencing attitudes and changing behavior.* Reading: Addison-Wesley, 1970.